



Bioinformática

Análisis de BRCA1

Entrega final

Ramiro Olivera Fedi

rolivera@itba.edu.ar

Julián Antonielli

jantonielli@itba.edu.ar

Introducción

El cáncer de mama afecta a una de cada ocho mujeres durante su vida¹. Es por esta razón que decidimos estudiar el gen `BRCA1`² (Breast Cancer 1), un gen supresor de tumores humano, que regula el ciclo celular y evita la proliferación incontrolada. En las mujeres portadoras de mutaciones en el gen `BRCA1`, el riesgo acumulado hasta los 70 años se estima entre 51 y 95% para cáncer de mama³.

Se seleccionó el mRNA `NM_007294`⁴ transcrito en isoforma 1 y se descargó en formato `GenBank` para el desarrollo y análisis del presente trabajo.

Implementación

El proyecto se desarrolló en Ruby con la gema `Bioruby` debido a la fluencia que tenemos en este lenguaje como desarrolladores. Se desarrollaron una serie de scripts, en particular, `Ex1.rb`, `Ex2.rb`, `Ex4.rb` y `Ex5.rb`, cuya API e instrucciones de ejecución se exponen en el archivo `README.md`.

Ejercicio 1

El primer ejercicio consiste en leer una secuencia de nucleótidos del mRNA seleccionado (`NM_007294`) del gen `BRCA1` en formato `GenBank` y traducirlos a las posibles secuencias de aminoácidos (Dependiendo del *reading frame*), finalmente guardandolos en distintos archivos de formato `FASTA`.

La idea de la implementación es leer el archivo de `GenBank` y simplemente iterar sobre los posibles *reading frames* (De 1 a 6) traduciendo las secuencias a aminoácidos utilizando la función librería `Bioruby`.

La complejidad de nuestra solución se remonta básicamente al *one-liner*

```
record.to_biobase.translate(frame)
```

que dada la representación del archivo `GenBank` de `Bioruby`, la transforma a una representación abstracta de una secuencia, y finalmente la traduce a la cadena de aminoácidos correspondientes con el frame indicado.

El script crea 6 nuevos archivos (uno por cada *reading frame*) con una posible secuencia de aminoácidos en el formato `FASTA`.

Ejercicio 2

Para el segundo ejercicio se tomaron las secuencias de aminoácidos generados en el ejercicio anterior en formato `FASTA` y se generó el script `Ex2.rb` que itera sobre los archivos dado el nombre del mRNA y hace consultas a `BLAST` remoto. Los resultados para

cada consulta se guardan en diferentes archivos con el mismo nombre que los generados en el ejercicio anterior pero con extensión `blas`.

El reporte generado por el `BLAST` incluye muchos resultados de interés. Si bien algunos son autodescriptivos, como el *Length of Overlapping region* que informa la cantidad de aminoácidos coincidentes entre la secuencia de la query y las encontradas en la base de datos, otros requieren de mayor atención como el valor estadístico *bit score*⁶.

Para evaluar si un alineamiento dado constituye evidencia de homología, es necesario entender que el alineamiento puede darse por simple casualidad. Es por esto que es necesario cuantificar de alguna manera la significancia de la alineación con respecto al puro azar. Esto se logra a través de los puntajes, o *scores*.

Por suerte, la estadística para el cálculo de puntajes en alineamientos locales está bien estudiada y definida. Esto es particularmente cierto para alineamientos locales carentes de huecos.

Estos alineamientos locales consisten de un par de segmentos de igual longitud. Si los puntajes para estos no pueden mejorarse por extensión o recortado se los conoce como *HSP*.

Para analizar qué tan probable es de que surja un puntaje alto por pura casualidad, se utiliza un modelo aleatorio de secuencias. En el caso particular de secuencias suficientemente largas con longitudes m y n , la estadística de los puntajes de *HSP* se caracterizan por el parámetro K y λ . En resumen, la cantidad de *HSPs* con puntaje de al menos S está dado por la siguiente fórmula conocida como *valor E para el puntaje S*:

$$E = Kmn e^{-\lambda S} \quad (1)$$

Trabajando un poco más la fórmula anterior a través de la normalización, podemos llegar a la siguiente fórmula, con unidades fijas:

$$S' = \frac{\lambda S - \ln K}{\ln 2} \quad (2)$$

Entonces el *valor E* correspondiente a un *bit score* dado es simplemente:

$$E = mn 2^{-S'} \quad (3)$$

Los *bit score* subsumen la esencia estadística del sistema de puntuación empleado, de modo que para calcular la significancia uno necesita saber sólo el tamaño del espacio de búsqueda.

Para reconocer qué *reading frame* es el correcto utilizamos dos estrategias: A partir de los resultados del BLAST y utilizando la frecuencia de *codones-stop*.

Comenzamos por la segunda. La traducción del mRNA comienza cuando el ribosoma encuentra la metionina y a partir de ahí comienza a traducir. En las regiones donde los *codones-stop* son muy frecuentes, en general, no puede traducirse ningún polipéptido que tenga alguna funcionalidad específica dada la baja cantidad de aminoácidos por cadena proteica producida. En el *reading frame* 2, hay una región de alrededor de 1800 aminoácidos consecutivos a la cual se le podría llegar a asignar una funcionalidad. En contraste, los otros *reading frames* se encuentran interrumpidos en la totalidad de la secuencia, es decir, con alta frecuencia de *codones-stop* a lo largo de la secuencia.

Si nos basamos en los resultados del BLAST es fácil notar que el *reading frame* correcto es el segundo. Solo el hecho que el *bit score* del primer resultado de la query hecha con ese *reading frame* sea de más de 3500, mientras que el siguiente mejor con otro *reading frame* es de 114 es evidencia suficiente para probarlo.

Ejercicio 3

Para realizar el alineamiento múltiple se utilizó la herramienta MUSCLE⁵, una herramienta de alineamiento de secuencias de proteínas y nucleótidos perteneciente al dominio público. MUSCLE es una de las herramientas de alineamiento múltiple con mejor performance de acuerdo a pruebas de referencias, con precisión y velocidad consistentemente mejores que CLUSTALW, con la capacidad de alinear cientos de secuencias en segundos.

Con los resultados obtenidos del BLAST de la secuencia de aminoácidos con el *reference frame* correcto, se seleccionaron 3 especies a comparar: *Pan paniscus*, *Pan troglodytes* y *Pongo abelii*.

Especie	mRNA
Homo Sapiens	NM_007294
Pan paniscus	NP_001288687
Pan troglodytes	XP_016785970
Pongo abelii	XP_003778880

Tabla 1 - mRNA seleccionado de cada especie

El repositorio de código contiene el archivo FASTA.mix con la concatenación de las 4 secuencias de aminoácidos, el output de la herramienta en el archivo MUSCLE.out y el resultado del alineamiento en alignment.html.



Figura 1 - Representación del alineamiento múltiple de secuencias de aminoácidos en MUSCLE

Analizando los resultados obtenidos del alineamiento podemos realizar algunas conclusiones con cierto grado de seguridad. Todas las secuencias coinciden en gran parte. De manera excepcional la cadena de aminoácidos del *Homo Sapiens* excede en el comienzo y en el final la de las especies alternativas. Podemos notar además que la secuencia conservada de la secuencia del *Homo Sapiens* (donde no hay *codones-stop*) es muy similar a la compartida entre todas las especies.

Dado que los genes en las distintas especies funcionan de igual manera, podemos concluir que la estructura de la proteína es la misma (o muy similar).

El hecho de que se trate de distintas especies podría implicar que los mismos resultados podrían obtenerse con proteínas estructuradas de distinta forma. No obstante, al tratarse de un gen regulador del ciclo celular, se trata de un mecanismo muy común conservado a lo largo de las distintas especies. Aún más en especies similares como las elegidas.

Finalmente, podemos concluir que los diferentes aminoácidos presentes en la mismas ubicaciones no alteran la estructura de la proteína, es decir, las mutaciones de estos fueron inocuas a su correcto funcionamiento.

Ejercicio 4

Se desarrolló el script `Ex4.rb` que dado un reporte de BLAST (como el obtenido en el segundo ejercicio) y un patrón, busca posibles coincidencias del patrón entre los *hits* del reporte de BLAST. Además se parsea el *accession* del hit seleccionado (donde hay una coincidencia del patrón) y con el módulo `Bio::DB::GenBank` se obtiene la secuencia completa del hit en formato FASTA y se escribe en el archivo de salida.

Se incluyeron además las opciones `--protein` y `--nucleotid` para elegir el tipo de BLAST.

Ejercicio 5

Para el quinto ejercicio se descargó EMBOSS, la base de datos de Prosite de sus fuentes oficiales y se generó el script `Ex5.rb`. Este script realiza dos funciones:

Por un lado, con la opción `--orf`, genera las cadenas de aminoácidos posibles dado el archivo GenBank obtenido en el primer ejercicio utilizando la función de EMBOSS `getorf`. Por el otro lado, con la opción `--prosite`, genera el análisis de dominios de una secuencia de aminoácidos en formato FASTA. Ésta última función se realiza utilizando por un lado la función `prosextract`, para preparar la base de Prosite, y la función `patmatmotifs`, que lee la secuencia de proteínas y la busca contra la base de datos Prosite de *motifs*. Finalmente genera un reporte EMBOSS estándar con detalles de la ubicación y el puntaje de cualquiera de los *motifs* que matcheen.

Ejercicio 6

Cómo se explicó en la introducción, se eligió el gen BRCA1 (<https://www.ncbi.nlm.nih.gov/gene/672>) ya que las mutaciones en este gen son responsables de aproximadamente 40% de los cánceres de mama heredados, y más del 80% entre cáncer de mama y ovario heredados.

BRCA1 es un gen que codifica una fosfoproteína que tiene un papel en el mantenimiento de la estabilidad genómica, actúa como un supresor tumoral, y juega un papel en la transcripción, reparación del ADN y su recombinación.

Se realizó una búsqueda de genes homólogos en distintas bases de datos. La base de datos de HomoloGene muestra el conjunto de genes identificados como homólogos del BRCA1 como se muestra en la tabla a continuación

Genes

Genes identified as putative homologs of one another during the construction of HomoloGene.

BRCA1, *H.sapiens*
breast cancer 1, early onset
BRCA1, *P.troglodytes*
breast cancer 1, early onset
BRCA1, *M.mulatta*
breast cancer 1, early onset
BRCA1, *C.lupus*
breast cancer 1, early onset
BRCA1, *B.taurus*
breast cancer 1, early onset
Brca1, *M.musculus*
breast cancer 1
Brca1, *R.norvegicus*
breast cancer 1, early onset
BRCA1, *G.gallus*
breast cancer 1, early onset

Figura 2 - Genes homólogos a BRCA1 como se muestran en HomoloGene

A partir de estos datos podemos ver que en principio existen 8 genes homólogos en la base de datos.

Continuamos con la base de `Ensembl` en la cual se obtuvieron 103 genes homólogos, visualizados a través del Gene Tree, en el cual se puede ver con más detalle las especies y las familias en la cuales está identificado algún gen homólogo al BRCA1.

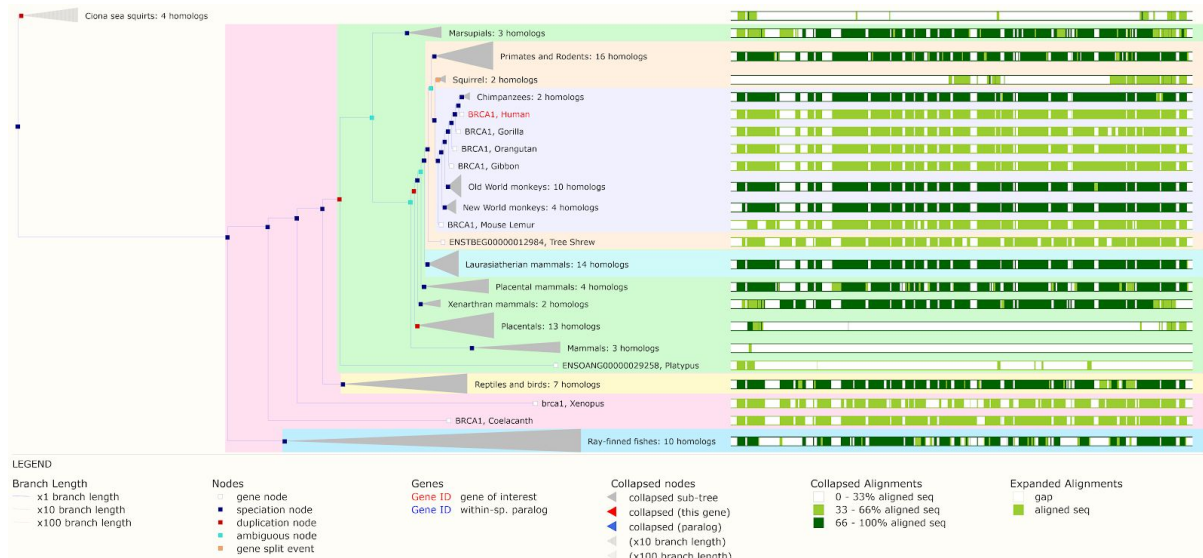


Figura 3 - Genes homólogos a BRCA1 como se muestran en el Gene Tree de Ensembl

Dado que se trata de un gen que genera supresores de tumores, y altamente relacionado con la reproducción celular, entendemos que es un gen que se encontrará en especies variadas, incluso de distintas familias. Por eso podemos ver que se encuentran desde mamíferos hasta reptiles y marsupiales.

En búsqueda de la distinta cantidad de transcritos y formas alternativas de splicing para BRCA1, se realizó una búsqueda en `Ensembl`, en el que se encontraron los siguientes 33 transcritos, como se describen en la siguiente figura.

Show/hide columns										Filter		
Name	Transcript ID	bp	Protein	Translation ID	Biotype	CCDS	UniProt	RefSeq	Flags			
BRCA1-210	ENST00000471181.7	7270	1884aa	ENSP00000418960	Protein coding	CCDS11456	P38398	NM_007300 NP_009231	TSL:1	GENCODE basic	APPRIS P4	
BRCA1-203	ENST00000357654.8	7207	1863aa	ENSP00000350283	Protein coding	CCDS11453	P38398	NM_007294 NP_009225	TSL:1	GENCODE basic	APPRIS ALT2	
BRCA1-221	ENST00000493795.5	5732	1816aa	ENSP00000418775	Protein coding	CCDS11459	P38398	NM_007297 NP_009228	TSL:5	GENCODE basic		
BRCA1-208	ENST00000468300.5	3273	699aa	ENSP00000417148	Protein coding	CCDS11455	P38398	NM_007299 NP_009230	TSL:1	GENCODE basic		
BRCA1-219	ENST00000491747.6	2379	759aa	ENSP00000420705	Protein coding	CCDS11454	A0A024R1V0 P38398	NM_007298 NP_009229	TSL:5	GENCODE basic		
BRCA1-202	ENST00000354071.7	4497	1399aa	ENSP00000326002	Protein coding	-	Q5YL82	-	TSL:1	GENCODE basic		
BRCA1-201	ENST00000352993.7	3668	721aa	ENSP00000312236	Protein coding	-	A0A024R1Z8 P38398	-	TSL:5	GENCODE basic		
BRCA1-232	ENST00000644379.1	2571	659aa	ENSP00000496570	Protein coding	-	-	-	CDS 5' incomplete			
BRCA1-230	ENST00000634433.1	2534	798aa	ENSP00000489431	Protein coding	-	A0A0U1RRA9	-	CDS 3' incomplete	TSL:5		
BRCA1-209	ENST00000470026.5	2108	649aa	ENSP00000419274	Protein coding	-	E7EWN5	-	CDS 3' incomplete	TSL:1		
BRCA1-214	ENST00000477152.5	1980	622aa	ENSP00000419988	Protein coding	-	E9PH68	-	CDS 3' incomplete	TSL:1		
BRCA1-215	ENST00000478531.5	1972	623aa	ENSP00000420412	Protein coding	-	E7EUM2	-	CDS 3' incomplete	TSL:1		
BRCA1-222	ENST00000493919.5	1948	572aa	ENSP00000418819	Protein coding	-	B7ZA85	-	CDS 3' incomplete	TSL:2		
BRCA1-223	ENST00000494123.5	1612	473aa	ENSP00000419103	Protein coding	-	Q3B891	-	CDS 3' incomplete	TSL:1		
BRCA1-216	ENST00000484087.5	1495	498aa	ENSP00000419481	Protein coding	-	H0Y8B8	-	CDS 5' and 3' incomplete	TSL:5		
BRCA1-226	ENST00000591534.5	1282	354aa	ENSP00000467329	Protein coding	-	K7EPC7	-	TSL:1	GENCODE basic		
BRCA1-212	ENST00000473961.5	958	319aa	ENSP00000420201	Protein coding	-	H0Y8D8	-	CDS 5' and 3' incomplete	TSL:2		
BRCA1-217	ENST00000487825.5	800	266aa	ENSP00000418212	Protein coding	-	H0Y881	-	CDS 5' and 3' incomplete	TSL:5		
BRCA1-225	ENST00000586385.5	781	173aa	ENSP00000465818	Protein coding	-	C6YB45	-	TSL:1	GENCODE basic		
BRCA1-224	ENST00000497488.1	779	177aa	ENSP00000418986	Protein coding	-	C9LZW4	-	CDS 3' incomplete	TSL:1		
BRCA1-213	ENST00000476777.5	769	222aa	ENSP00000417554	Protein coding	-	E7EQW4	-	CDS 3' incomplete	TSL:5		
BRCA1-206	ENST00000461574.1	726	242aa	ENSP00000417241	Protein coding	-	H0Y850	-	CDS 5' and 3' incomplete	TSL:2		
BRCA1-228	ENST00000618469.1	718	23aa	ENSP00000478114	Protein coding	-	G1UI37	-	CDS 3' incomplete	TSL:1		
BRCA1-227	ENST00000591849.5	563	96aa	ENSP00000465347	Protein coding	-	K7EJW3	-	TSL:1	GENCODE basic		
BRCA1-233	ENST00000644555.1	558	65aa	ENSP00000494614	Protein coding	-	-	-	CDS 3' incomplete			
BRCA1-218	ENST00000489037.1	455	98aa	ENSP00000420781	Protein coding	-	E7ENB7	-	CDS 3' incomplete	TSL:4		
BRCA1-204	ENST00000412061.3	1312	437aa	ENSP00000397145	Non stop decay	-	A0A0A0MSN1	-	CDS 5' incomplete	TSL:1		
BRCA1-205	ENST00000461221.5	5693	63aa	ENSP00000418548	Nonsense mediated decay	-	P38398	NR_027676	TSL:1			
BRCA1-220	ENST00000492859.5	1584	59aa	ENSP00000420253	Nonsense mediated decay	-	E9PC22	-	TSL:1			
BRCA1-231	ENST00000642945.1	1255	81aa	ENSP00000495897	Nonsense mediated decay	-	-	-				
BRCA1-207	ENST00000461798.5	582	63aa	ENSP00000417988	Nonsense mediated decay	-	P38398	-	TSL:3			
BRCA1-229	ENST00000621897.1	375	No protein	-	Processed transcript	-	-	-	TSL:5			
BRCA1-211	ENST00000472490.1	561	No protein	-	Retained intron	-	-	-	TSL:3			

Figura 4 - Transcritos de BRCA1

De las variantes de splicing que se exponen en la figura anterior, solo se expresan 26. Las diferencias entre los tamaños de las proteínas nos sugiere que la estructura será distinta y por lo tanto sus respectivas funciones diferentes. Además, al estar identificadas por distintos nombres, se nos sugiere que están involucradas en distintos mecanismos.

Por la cantidad y calidad de resultados, asumimos que Ensembl cuenta con una base de datos más rica en lo que respecta a transcritos.

Se busca ahora analizar las interacciones entre la proteína codificada por BRCA1 y distintas proteínas.

Utilizando NCBI podemos encontrar 1150 proteínas que interactúan con el gen BRCA1, y utilizando Uniprot, encontramos que la base de datos BioGrid logra identificar alrededor de 1000, como se muestra en la siguiente figura.

Protein-protein interaction databases

BioGrid ⁱ	107140, 1004 interactors
CORUM ⁱ	P38398
DIP ⁱ	DIP-5971N
IntAct ⁱ	P38398, 75 interactors
MINT ⁱ	P38398
STRING ⁱ	9606.ENSP00000418960

Figura 5 - Interacciones entre distintas proteínas y BRCA1

Utilizando las herramientas de visualización de BioGrid se armó un grafo de las interacciones entre las proteínas⁷.

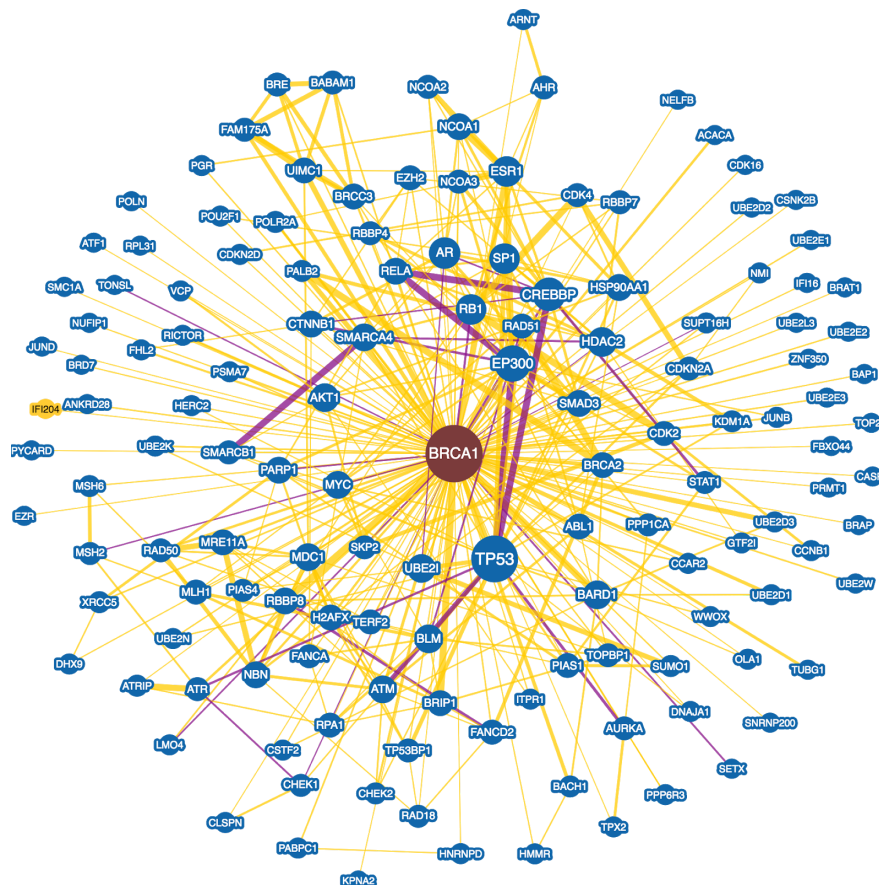


Figura 6 - Interacciones entre distintas proteínas y BRCA1

Debido a la gran cantidad de interacciones no es fácil identificar un patrón entre las mismas. No obstante, analizando algunas de las proteínas que interactúan con BRCA1 podemos notar que la mayoría están asociadas a mecanismos de reparación o identificación de daño genético. Lo que nos lleva a confirmar que dicha proteína está asociada a *pathways* de reparación genética.

Analizando un caso particular, podemos ver la interacción de BRCA1 con la proteína RAD51. Ésta última se encarga de asistir en la reparación de daños de cadena doble en el ADN, y

puede encontrarse en la mayoría de las células eucariotas, desde levadura hasta células humanas⁸.

En cuanto a la relación que existe entre *BRCA1* y *RAD51*, encontramos que la primera regula la segunda en respuesta al daño en el ADN con implicancias en la cohesión de cromátidas hermanas, estabilidad genómica y en la carcinogénesis⁹

Utilizando la base de datos de genes humanos *GeneCards*, buscamos información sobre los procesos biológicos a los que pertenece *BRCA1*, las funciones moleculares en las que trabaja y los componentes celulares que forman parte de la misma¹⁰.

Como puede verse en la próxima figura, se encontraron los siguientes 16 componentes celulares

GO ID	Qualified GO term	Evidence	PubMed IDs
GO:0000151	ubiquitin ligase complex	NAS	14976165
GO:0000793	condensed chromosome	IEA	
GO:0000794	condensed nuclear chromosome	IEA	
GO:0000800	lateral element	IDA	9774970
GO:0005634	nucleus	IEA,IDA	17525340
GO:0005654	nucleoplasm	TAS	
GO:0005694	chromosome	ISS	
GO:0005737	cytoplasm	IEA,IDA	20160719
GO:0005886	plasma membrane	IDA	21282464
GO:0008274	gamma-tubulin ring complex	NAS	12214252
GO:0016020	membrane	IEA	
GO:0016021	integral component of membrane	IEA	
GO:0030529	intracellular ribonucleoprotein complex	IDA	18809582
GO:0031436	BRCA1-BARD1 complex	IDA	12890688
GO:0043234	protein complex	IDA	9774970
GO:0070531	BRCA1-A complex	IDA	17525340

Figura 7 - Componentes celulares de BRCA1

Utilizando el mismo sitio encontramos 68 procesos biológicos a los que pertenece *BRCA1*. En la siguiente figura se muestran solo algunos de ellos.

GO ID	Qualified GO term	Evidence	PubMed IDs
GO:0000724	double-strand break repair via homologous recombination	IDA	17349954
GO:0000729	DNA double-strand break processing	TAS	
GO:0000731	DNA synthesis involved in DNA repair	TAS	
GO:0000732	strand displacement	TAS	
GO:0006260	DNA replication	TAS	
GO:0006281	DNA repair	IEA	
GO:0006301	postreplication repair	IDA	17349954
GO:0006302	double-strand break repair	IMP,IDA	17525340
GO:0006303	double-strand break repair via nonhomologous end joining	TAS	
GO:0006310	DNA recombination	IEA	
GO:0006349	regulation of gene expression by genetic imprinting	IEA	
GO:0006351	transcription, DNA-templated	IEA	
GO:0006355	regulation of transcription, DNA-templated	IEA	
GO:0006357	regulation of transcription by RNA polymerase II	TAS,IMP	9662397
GO:0006359	regulation of transcription by RNA polymerase III	TAS	10918303
GO:0006366	transcription by RNA polymerase II	IEA	

Figura 8 - Procesos biológicos a los que pertenece BRCA1

Continuando con el análisis de los resultados, encontramos que esta proteína trabaja en las siguientes 16 funciones moleculares

GO ID	Qualified GO term	Evidence	PubMed IDs
GO:0001105	RNA polymerase II transcription coactivator activity	IMP	9662397
GO:0003677	DNA binding	TAS,IEA	9662397
GO:0003684	damaged DNA binding	IEA	
GO:0003713	transcription coactivator activity	NAS	15572661
GO:0003723	RNA binding	IDA	12419249
GO:0004842	ubiquitin-protein transferase activity	IDA,IEA	12890688
GO:0005515	protein binding	IPI	8944023
GO:0008270	zinc ion binding	IEA	
GO:0015631	tubulin binding	NAS	12214252
GO:0016740	transferase activity	IEA	
GO:0019899	enzyme binding	IPI	15965487
GO:0031625	ubiquitin protein ligase binding	IPI	17873885
GO:0044212	transcription regulatory region DNA binding	IDA	20820192
GO:0046872	metal ion binding	IEA	
GO:0050681	androgen receptor binding	NAS	15572661
GO:0070063	RNA polymerase binding	IDA	9662397

Figura 9 - Funciones moleculares en las que trabaja BRCA1

Buscamos ahora analizar las vías metabólicas en las que participa *BRCA1*. De acuerdo a Reactome, participa de 5 grandes funciones de alto nivel:

- Ciclo celular
- Reparación de ADN
- Expresión génica
- Metabolismo de proteínas
- Reproducción

En la siguiente figura podemos ver algunos pathways específicos en los que participa *BRCA1*, asociados a la expresión génica.

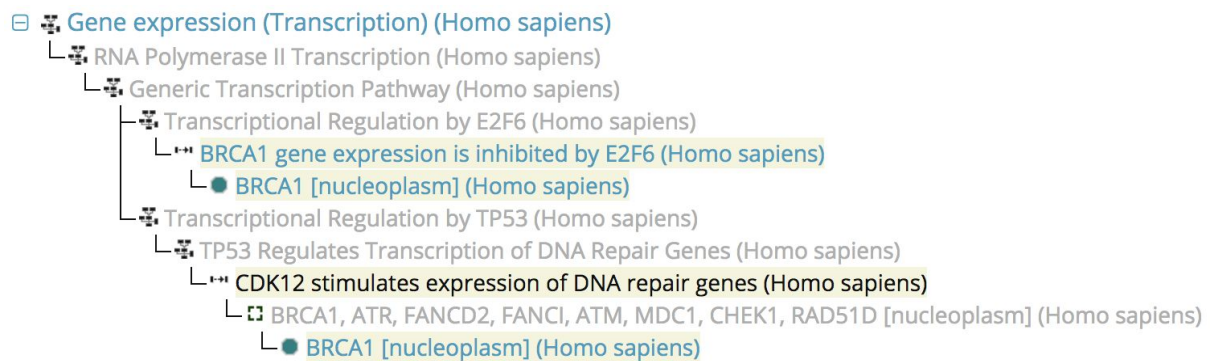


Figura 10 - Pathways en los que participa *BRCA1* asociados a la expresión génica

Analizamos el caso particular de cómo interviene *BRCA1* en la vía metabólica de *CDK12*, que se encarga de estimular la expresión de genes reparadores de ADN.

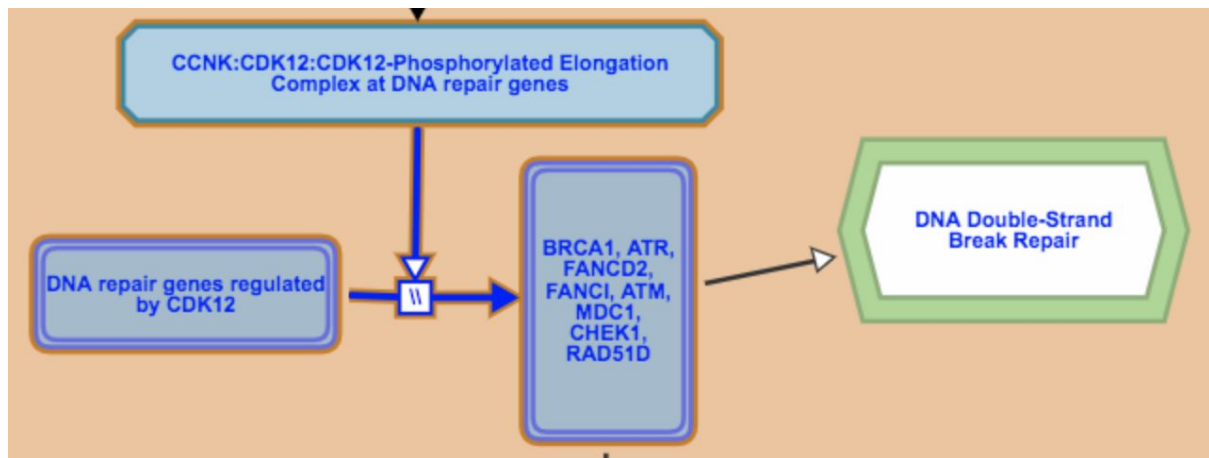


Figura 11 - Mecanismo de la vía metabólica de *CDK12* y cómo interviene *BRCA1*

Se probó que *CDK12* colocaliza con el complejo de ARN Polimerasa II (Catalizador de la transcripción del ADN) en *FANCD2*, *FANCI*, *ATM*, *CHEK1*, *MDC1*, *RAD51D* y *ATR*, además de ser necesario para alcanzar una expresión suficiente de *BRCA1*. Por otro lado, mutaciones

en CDK12 se encontraron de manera recurrentes en tumores de cáncer de ovario. Dichas mutaciones afectan el efecto catalítico de CDK12 u ocasionan la pérdida total de las funciones del gen. Los tumores de ovarios que albergan mutaciones que inactivan CDK12 exhiben niveles decrecidos de BRCA1¹¹.

Como se discutió hasta ahora BRCA1 es un gen supresor de tumores. Como la mayoría de los genes, las variaciones de BRCA1 pueden ser causales para determinada enfermedad (El cáncer de mama en este caso), asociadas con un riesgo mayor, o benignos.

De todas formas, *causal* no significa que exista un 100% de seguridad de que la persona con tal variante vaya a desarrollar la enfermedad. Para los individuos con variantes de BRCA1 causales por ejemplo, la OMIM¹² resume los siguientes valores:

- El riesgo de cáncer de mama a lo largo de la vida para portadores de mutaciones es de entre el 80% y el 90%
- El riesgo de cáncer de ovario a lo largo de la vida para portadores de mutaciones es de entre el 40% y el 50%
- Aumenta el riesgo de cáncer de mama bilateral

De manera similar, una publicación del 2007 que involucró cerca de 10.000 mujeres a lo largo de 20 años concluyó que las mujeres con mutaciones en BRCA1 tienen en promedio 72% de riesgo de desarrollar cáncer mamario para la edad de 80 años, y que el promedio del riesgo de desarrollar cáncer de ovario a lo largo de su vida es del 44%¹³

Hay más de 500 variaciones de BRCA1 que son consideradas causales, sin embargo casi todas ellas son muy difíciles de encontrar. Se estima que todas las mutaciones de BRCA1 causales combinadas ocurren en menos de 0.3% de la gente.

Algunas de las mutaciones menos raras, causales del cáncer de mama son rs386833395¹⁴ y rs80357906¹⁵, mejor conocidas como 185delAG y 5382insC.

Dichas variantes afectan principalmente al grupo étnico conocido como Judíos Asquenazí o asquenazi. Esta denominación se refiere a los judíos que se asentaron en Europa Central y Oriental. Se establecieron principalmente en Alemania, Austria, Hungría, República Checa, Eslovaquia, Polonia, Ucrania, Rumania, Moldavia, Rusia, Bielorrusia, Bulgaria, Lituania y Letonia. Los asquenazíes son los descendientes de las comunidades judías medievales establecidas a lo largo del Rin, desde Alsacia, al sur, hasta Renania, en el norte¹⁶.

Otras variantes conocidas, con mayor o menor consecuencias se encuentran en la siguiente tabla¹⁷.

Código	Nombre
rs28897696	A1708E

rs55770810	R1699W
rs1799950	Q356R
rs4986850	D693N
rs2227945	S1140G
rs16942	K1183R
rs1799966	S1613G
rs41293463	M1775R

Tabla 2 - Variaciones de BRCA1

Referencias

1. <https://medlineplus.gov/spanish/breastcancer.html>
2. BREAST CANCER 1 GENE; BRCA1. <https://www.omim.org/entry/113705>
3. Efectividad de los protocolos de prevención en mujeres portadoras de las mutaciones BRCA1/2.
<http://www.update-software.com/BCP/BCPGetDocument.asp?DocumentID=GCS33-7>
4. Homo sapiens BRCA1, DNA repair associated (BRCA1), transcript variant 1, mRNA.
https://www.ncbi.nlm.nih.gov/nuccore/NM_007294
5. Edgar RC (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". Nucleic Acids Research.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337>
6. The Statistics of Sequence Similarity Scores.
<https://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>
7. <https://thebiogrid.org/107140>
8. Shinohara A, Ogawa H, Ogawa T (May 1992). "Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein". Cell. 69 (3): 457–70.
doi:10.1016/0092-8674(92)90447-K. PMID 1581961.
9. <https://www.ncbi.nlm.nih.gov/pubmed/16357146>
10. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=BRCA1&keywords=BRCA1>
11. <https://reactome.org/PathwayBrowser/#/R-HSA-6796648&SEL=R-HSA-6797712&PATH=R-HSA-74160,R-HSA-73857,R-HSA-212436,R-HSA-3700989>
12. <http://www.omim.org/clinicalSynopsis/604370>
13. <https://www.ncbi.nlm.nih.gov/pubmed/28632866?dopt=Abstract>
14. https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=386833395
15. https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=80357906
16. Joan Comay, The Diaspora Story, Tel Aviv y Jerusalén: Steimatzky, 1981, pp. 138-39.
17. <https://www.snpedia.com/index.php/BRCA1>